

HOUSE PRICE INDIA

NAME : M SIVASAKTHI

Untitled13.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

Files

- sample_data
 - README.md
 - anscombe.json
 - california_housing_test.csv
 - california_housing_train.csv
 - mnist_test.csv
 - mnist_train_small.csv
 - House Price India.csv
 - archive.zip

Code

```
[1] !unzip '/content/archive.zip'
```

Archive: /content/archive.zip
replace House Price India.csv? [y]es, [n]o, [A]ll, [N]one, [r]ename: y
inflating: House Price India.csv

```
[2] import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```

```
[3] df=pd.read_csv('/content/House Price India.csv')
```

```
[4] df.head()
```

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year	Pos C
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122

completed at 8:51 AM

Untitled13.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

Files

- sample_data
 - README.md
 - anscombe.json
 - california_housing_test.csv
 - california_housing_train.csv
 - mnist_test.csv
 - mnist_train_small.csv
 - House Price India.csv
 - archive.zip

Code

```
2 6762810998 42491 5 2.75 2910 9480 1.5 0 0 3 ... 1939 0 122  
3 6762812605 42491 4 2.50 3310 42998 2.0 0 0 3 ... 2001 0 122  
4 6762812919 42491 3 2.00 2710 4500 1.5 0 0 4 ... 1929 0 122
```

5 rows x 23 columns

```
[5] df.shape
```

(14620, 23)

```
[6] ##Univariate Analysis  
df_price = df.loc[df['Price']>=3000000]  
df_year = df.loc[df['Built Year']>1990]  
df_ryear = df.loc[df['Renovation Year']>2000]
```

```
[7] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 14620 entries, 0 to 14619  
Data columns (total 23 columns):  
# Column Non-Null Count Dtype  
---  
0 id 14620 non-null int64  
1 Date 14620 non-null int64
```

completed at 8:51 AM

Untitled13.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings

Files

sample_data

README.md

anscombe.json

california_housing_test.csv

california_housing_train.csv

mnist_test.csv

mnist_train_small.csv

House Price India.csv

archive.zip

+ Code + Text

0s Data columns (total 23 columns):

#	Column	Non-Null Count	Dtype
0	id	14620 non-null	int64
1	Date	14620 non-null	int64
2	number of bedrooms	14620 non-null	int64
3	number of bathrooms	14620 non-null	float64
4	living area	14620 non-null	int64
5	lot area	14620 non-null	int64
6	number of floors	14620 non-null	float64
7	waterfront present	14620 non-null	int64
8	number of views	14620 non-null	int64
9	condition of the house	14620 non-null	int64
10	grade of the house	14620 non-null	int64
11	Area of the house(excluding basement)	14620 non-null	int64
12	Area of the basement	14620 non-null	int64
13	Built Year	14620 non-null	int64
14	Renovation Year	14620 non-null	int64
15	Postal Code	14620 non-null	int64
16	Lattitude	14620 non-null	float64
17	Longitude	14620 non-null	float64
18	living_area_renov	14620 non-null	int64
19	lot_area_renov	14620 non-null	int64
20	Number of schools nearby	14620 non-null	int64
21	Distance from the airport	14620 non-null	int64
22	Price	14620 non-null	int64

dtypes: float64(4), int64(19)
memory usage: 2.6 MB

0s completed at 8:51 AM

Untitled13.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings

Files

sample_data

README.md

anscombe.json

california_housing_test.csv

california_housing_train.csv

mnist_test.csv

mnist_train_small.csv

House Price India.csv

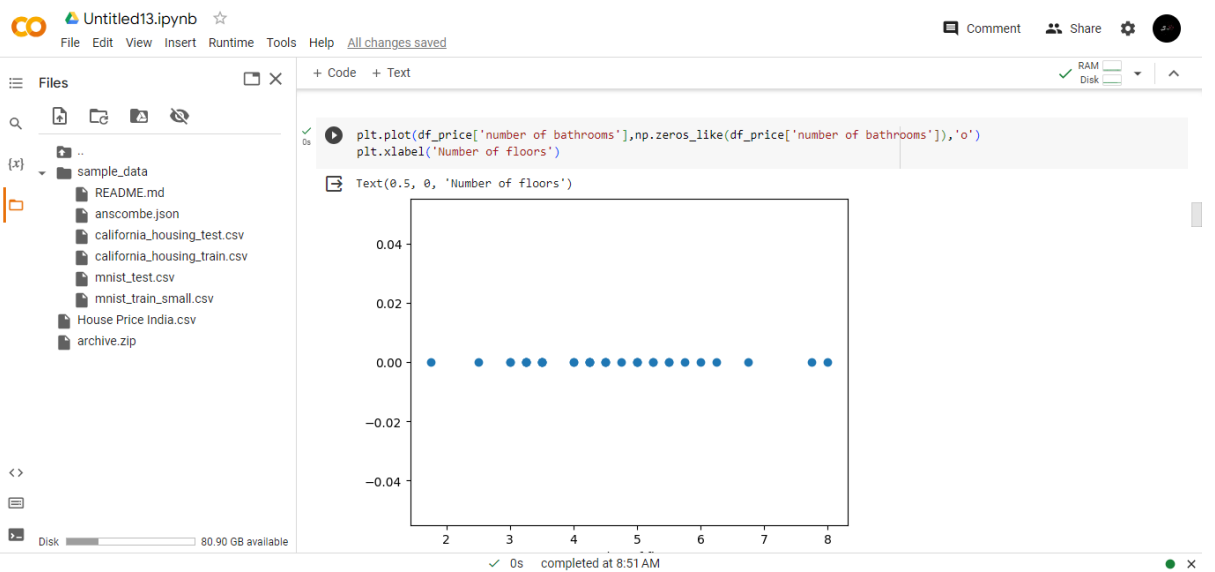
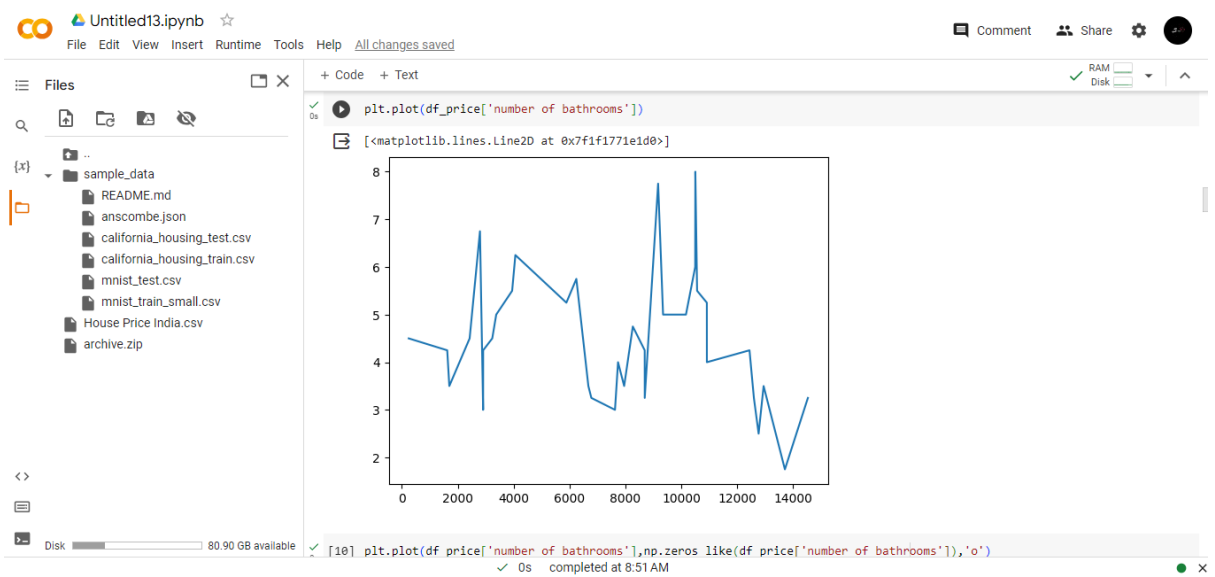
archive.zip

+ Code + Text

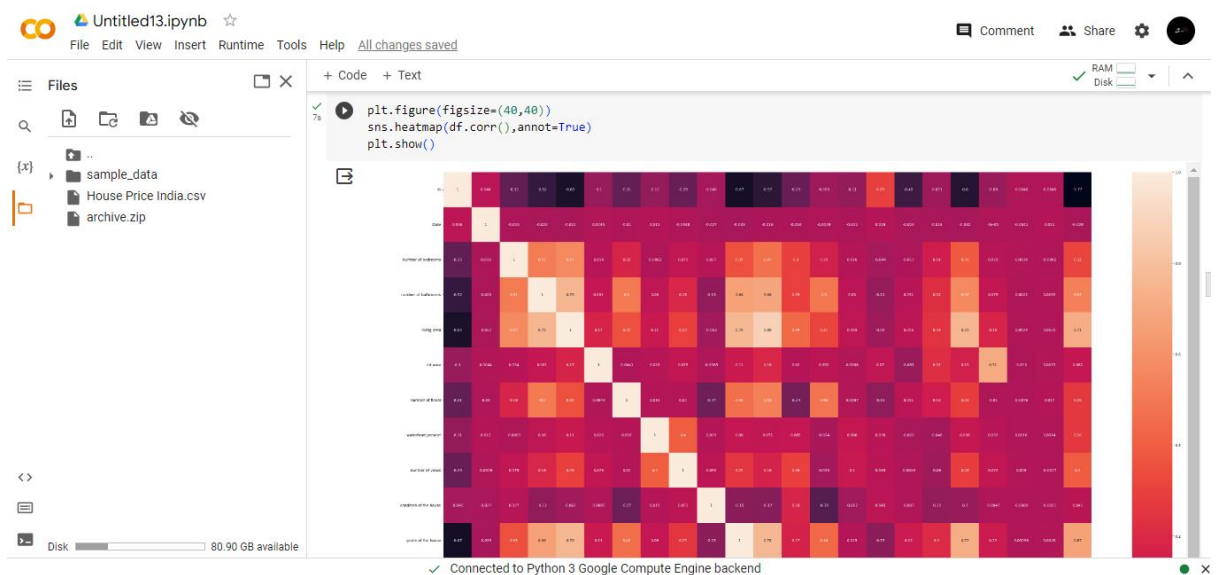
0s df_price

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year
243	6762810052	42496	7	4.50	6210	8856	2.5	0	2	5	...	1910	
1622	6762810059	42518	6	4.25	6980	15682	3.0	0	4	4	...	1999	
1697	6762810035	42519	4	3.50	5550	28078	2.0	0	2	4	...	2000	
2424	6762810021	42531	5	4.50	10040	37325	2.0	1	2	3	...	1940	2000
2794	6762810027	42537	5	6.75	9640	13068	1.0	1	4	3	...	1983	2000
2907	6762810029	42538	4	3.00	6430	27517	2.0	0	0	3	...	2001	
2908	6762810065	42538	4	4.25	4850	12445	2.0	1	4	5	...	1989	
3234	6762810043	42543	3	4.50	5230	17826	2.0	1	4	3	...	2005	
3376	6762810062	42544	4	5.00	4550	18641	1.0	1	4	3	...	2002	
3946	6762810033	42551	5	5.50	7050	42840	1.0	0	2	4	...	1978	
4061	6762810047	42552	5	6.25	8020	21738	2.0	0	0	3	...	2001	
5887	6762810060	42579	5	5.25	5090	23669	2.0	0	0	3	...	2006	

0s completed at 8:51 AM









Untitled13.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- House Price India.csv
- archive.zip

```
df.describe()
df['number of bathrooms'].value_counts()
```

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of units
count	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000	1.462000e+04	14620.000000	14620.000000	14620.000000
mean	6.762821e+09	42604.538646	3.379343	2.129583	2098.262996	1.509328e+04	1.502360	0.007661	0.250000
std	6.237575e+03	67.347991	0.938719	0.769934	928.275721	3.791962e+04	0.540239	0.087193	0.769934
min	6.762810e+09	42491.000000	1.000000	0.500000	370.000000	5.200000e+02	1.000000	0.000000	0.000000
25%	6.762815e+09	42546.000000	3.000000	1.750000	1440.000000	5.010750e+03	1.000000	0.000000	0.000000
50%	6.762821e+09	42600.000000	3.000000	2.250000	1930.000000	7.620000e+03	1.500000	0.000000	0.000000
75%	6.762826e+09	42662.000000	4.000000	2.500000	2570.000000	1.080000e+04	2.000000	0.000000	0.000000
max	6.762832e+09	42734.000000	33.000000	8.000000	13540.000000	1.074218e+06	3.500000	1.000000	4.000000

8 rows x 23 columns

```
[17] df['number of bathrooms'].value_counts()
```

number of bathrooms	count
1	~2.50e+03
2	~3.67e+03

Connected to Python 3 Google Compute Engine backend

Untitled13.ipynb ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

Comment Share Settings

Files

- sample_data
- House Price India.csv
- archive.zip

21s

```
df['number of bathrooms'].value_counts().to_frame()
```

number of bathrooms	
2.50	3678
1.00	2509
1.75	2062
2.25	1378
2.00	1323
1.50	968
2.75	831
3.00	510
3.50	504
3.25	424
3.75	101
4.00	81
4.50	72

80.90 GB available

Connected to Python 3 Google Compute Engine backend

Untitled13.ipynb ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

Comment Share Settings

Files

- sample_data
- House Price India.csv
- archive.zip

0s

```
df['number of bathrooms'].value_counts()
```

2.50	3678
1.00	2509
1.75	2062
2.25	1378
2.00	1323
1.50	968
2.75	831
3.00	510
3.50	504
3.25	424
3.75	101
4.00	81
4.50	72
4.25	56
0.75	47
4.75	17
5.00	15
5.25	12
5.50	8
1.25	7
0.50	3
6.00	3
8.00	2
5.75	2
6.25	2
6.75	2
6.50	1
7.50	1

80.90 GB available

Connected to Python 3 Google Compute Engine backend

Untitled13.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

Files

- sample_data
- House Price India.csv
- archive.zip

new_count = df['number of bathrooms'].value_counts().to_frame()
new_count.rename(columns={'number of bathrooms': 'new count'}, inplace = True)
new_count

	new count
2.50	3678
1.00	2509
1.75	2062
2.25	1378
2.00	1323
1.50	968
2.75	831
3.00	510
3.50	504
3.25	424
3.75	101
4.00	81
4.50	72

80.90 GB available

Connected to Python 3 Google Compute Engine backend

Untitled13.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

Files

- sample_data
- House Price India.csv
- archive.zip

new_count.index.name = "s.no"
new_count

s.no	new count
2.50	3678
1.00	2509
1.75	2062
2.25	1378
2.00	1323
1.50	968
2.75	831
3.00	510
3.50	504
3.25	424
3.75	101
4.00	81

80.90 GB available

Connected to Python 3 Google Compute Engine backend

Untitled13.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

Files

sample_data

House Price India.csv

archive.zip

<>

Disk 80.90 GB available

+ Code + Text

df.isnull()

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year	Postal Code
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False
1	False	False	False	False	False	False	False	False	False	False	...	False	False	False
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False
...
14615	False	False	False	False	False	False	False	False	False	False	...	False	False	False
14616	False	False	False	False	False	False	False	False	False	False	...	False	False	False
14617	False	False	False	False	False	False	False	False	False	False	...	False	False	False
14618	False	False	False	False	False	False	False	False	False	False	...	False	False	False
14619	False	False	False	False	False	False	False	False	False	False	...	False	False	False

14620 rows x 23 columns

Connected to Python 3 Google Compute Engine backend

Untitled13.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

Files

sample_data

House Price India.csv

archive.zip

<>

Disk 80.90 GB available

+ Code + Text

df.isnull().sum()

```
id 0
Date 0
number of bedrooms 0
number of bathrooms 0
living area 0
lot area 0
number of floors 0
waterfront present 0
number of views 0
condition of the house 0
grade of the house 0
Area of the house(excluding basement) 0
Area of the basement 0
Built Year 0
Renovation Year 0
Postal Code 0
Latitude 0
Longitude 0
living_area_renov 0
lot_area_renov 0
Number of schools nearby 0
Distance from the airport 0
Price 0
dtype: int64
```

Connected to Python 3 Google Compute Engine backend